Institute for Advanced Simulation

Multiscale Methods for the Description of Chemical Events in Biological Systems

Marcus Elstner and Qiang Cui

published in

Multiscale Simulation Methods in Molecular Sciences,
J. Grotendorst, N. Attig, S. Blügel, D. Marx (Eds.),
Institute for Advanced Simulation, Forschungszentrum Jülich,
NIC Series, Vol. **42**, ISBN 978-3-9810843-8-2, pp. 421-444, 2009.

http://www.fz-juelich.de/nic-series/volume42

Multiscale Methods for the Description of Chemical Events in Biological Systems

Marcus Elstner[1,2] **and Qiang Cui**[3]

[1] Department of Physical and Theoretical Chemistry
Technische Universität Braunschweig
D-38106 Braunschweig, Germany
[2] Department of Molecular Biophysics
German Cancer Research Center
D-69115 Heidelberg, Germany
E-mail: m.elstner@tu-bs.de

[3] Department of Chemistry and Theoretical Chemistry Institute
University of Wisconsin-Madison
1101 University Avenue, Madison, WI 53706, USA
E-mail: cui@chem.wisc.edu

Computational methods for the description of chemical events in biological structures have to take into account the key features of bio-molecular molecules, their high degree of structural flexibility and the long-range nature of electrostatic forces. In the last decade, a multitude of approaches have been developed to combine computational methods that span different length- and time-scales. These multiscale approaches incorporate a quantum mechanical description of the actives site in combination with an empirical force field method for the immediate protein environment and a continuum treatment of the regions further away. To study reactive events, efficient sampling techniques have to be applied, which can become computationally intense and therefore requires effective quantum methods. In this contribution, we describe the various options to combine different methods, where the specific combination depends very much on the nature of the problem in hand.

1 Introduction

The simulation of structure and dynamics of biological systems can nowadays be routinely performed using empirical force fields, which have become robust and reliable tools over the last decades[1,2]. These Molecular Mechanics (MM) force fields[3,4] model chemical bonds by harmonic springs, i.e. they describe the energy of a chemical bond using harmonic (or Fourier) potentials for the bond length, bond angle and dihedral angle. In addition to these bonded terms, the force fields contain non-bonded contributions, modeled by the interaction of fixed atomic point charges and van der Waals interactions, usually described by the 12-6 Lennard-Jones potential. Polarizable force fields[5] that allow the partial charges to vary depending on their environment have also been developed, although their applications have been much more limited due to the higher computational expense.

Biological structures host a multitude of chemical events like chemical reactions (biocatalysis), photochemical processes, long range proton transfers (e.g., in bioenergetics), electron and energy (excitation) transfers, which can only be described using quantum mechanical (QM) techniques and not with MM. The description of these processes is very challenging for computational chemistry due to the large size of biological systems and the presence of multiple time-scales. Indeed, biological structures take the middle ground

between solids and more disordered materials like polymers. On the one hand, they have a highly ordered structure from a functional perspective; e.g., specific functional amino acids with pre-organized orientations are found in the immediate vicinity of the active site, which is one important reason that chemical events in the enzyme active site are more efficient than the corresponding processes in solution[6]. On the other hand, biomolecules are highly flexible and entropic contributions to the reaction free energy can be as important as potential energy contributions. Therefore, to model chemical events in biological systems requires both accurate potential functions and access to sufficient conformational sampling and long time-scales.

None of the existing methods alone is up to the task in general. For example, standard QM methods like Hartree-Fock (HF), Density-Functional (DFT) or Semi-Empirical (SE) Theory alone can not handle several thousands of atoms with sufficient sampling. As a consequence, many studies in the past focused only on small parts of the system, such as the active site of the protein where the reaction occurs. This however, has been shown to be insufficient due to the long range nature of the electrostatic forces and steric interactions of the active site with the environment[6–8]. The development of linear scaling methods extended the applicability of QM significantly. However, their application to large systems is still costly, not viable for many interesting systems with 10,000-100,000 atoms and not helpful when dynamical or thermodynamical properties are required, which is the case in many biological applications. Evidently, methods from different computational levels have to be combined effectively, which has been explored for the past few decades.

In the quantum chemistry community, efforts have largely been focussed on the combination of QM methods with continuum electrostatic theories, starting from Born & Onsager theories that aimed at computing the solvation free energy of charges in a polar environment. These methods have been refined over the years and can now give a reasonable description of solvation properties in an isotropic and homogeneous medium[9,10]. In this context, MM force field methods have also been combined with continuum electrostatics methods[11,12] since the number of water that has to be included in explicit solvent simulations with the periodic boundary condition often far exceeds the number of atoms in the biological molecule itself. Most of these methods are based on the Poisson-Boltzmann theory[13] and the Generalized Born model[14], although more sophisticated integral equation and density functional theories[13] have also been explored for small biomolecules.

These continuum models (CM), however, are by no means appropriate to represent the electrostatic and steric interactions of the structured environment with the active site. Therefore, Warshel and Levitt[15] proposed in 1976 to combine QM methods for the active site with MM methods for the remainder of the system. An appropriate QM-MM coupling term describes the polarization of the QM region by the charges on the MM atoms and mediate the steric interactions via covalent bonds and van der Waals contacts. Up to now, such QM/MM methods have been developed to combine many QM methods (post-HF, HF, DFT, SE) with various force fields (e.g., CHARMM, AMBER, GROMOS, ...) and have become a powerful tool for analyzing chemical events in biomolecules.

It has long been envisioned that a multiscale model can be developed for complex molecular systems in which QM/MM methods are further augmented by a continuum electrostatic model. Indeed, although efficient Ewald summation has been implemented with QM/MM potential function[16,17], the high cost and sampling challenge associated with explicit solvent simulations also becomes more transparent for QM/MM simula-

tions, especially those using high level QM methods. Practical implementations that integrate QM/MM potential with continuum electrostatics models, however, only have become available in recent years[18–20]. The major focus of this review is to summarize the key components of such QM/MM/CM models and to discuss a few relevant examples that best illustrate their value and limitations.

2 QM/MM Methods

The development of QM/MM methods in recent years has turned them into powerful predictive tool and many research groups are involved in the process; most of the recent developments have been nicely summarized in a comprehensive review[21] (see the contribution of W. Thiel). There is not one single QM/MM method, and the multitude of different implementations can be characterized by several main distinctions:

- **Additive and subtractive methods:** Subtractive models[22] apply the QM method to the active site and the MM method to the entire system, also including the active site. Since the active site region is treated by both methods, the MM contribution for the active site has to be subtracted out:

$$E = E_{MM}^{tot} + E_{QM}^{active\ site} - E_{MM}^{active\ site} \tag{1}$$

  The advantage of this method is that it allows in a simple way to also combine two different QM methods in a QM/QM' scheme or multiple methods in a QM/QM'/MM scheme, where high (e.g., DFT) and low level (SE) QM methods are combined[23,24]. The disadvantage is that the MM has to treat also the active site, which may not be straightforward when the active site has complex electronic structure (e.g., transition metal centers). The additive scheme[25], by contrast, only applies the MM to the environment of the active site, and the two regions are then coupled by a QM/MM coupling term:

$$E = E_{QM}^{active\ site} + E_{MM}^{environement} + E_{QM/MM} \tag{2}$$

  Here, no force field parameters are needed for the active site, but the description of the boundary is conceptionally more involved.

- **The treatment of the QM/MM boundary:** In many applications, this boundary dissects a covalent bond. In the simplest *link atom* approach[25], the dangling bond of the QM region is saturated by an additional hydrogen. Other approaches avoid the introduction of this artificial hydrogen. The *pseudoatom/bond* approach[26] treats the frontier functional group as a pseudo-atom with an effective one-electron potential. In most cases, a C-C single bond has to be cut and the CH_2 at the QM boundary is then substituted by a parametrized (using a pseudo-potential) pseudo-Fluorine, which models the properties of the C-C bond. The hybrid-orbital approach[27] does not substitute the boundary CH_2 group but freezes the occupation of the orbital, which represents the dangling bond. These are the most common approaches to deal with the QM/MM boundary and various variants have also been proposed[28]. Systematic studies indicate that most schemes give comparable results as far as the charges at the QM/MM boundary are carefully treated[29–31].

- **Mechanical, electrostatic and polarizable embedding:** This concerns the QM/MM coupling term and the nature of the force field. In the mechanical embedding[22,23], the MM point charges are not allowed to polarize the QM region. The interaction of the QM and MM regions is simply given by the Coulomb and van der Waals interactions[32] between the QM and MM subsystems and the interactions at the boundary, thus the QM density is *not* perturbed by the MM charges. Since biological systems are often highly charged, this method should not be used for biological applications. The electrostatic embedding[25] includes the MM charges as external point charges in the determination of the QM charge density, i.e., the QM region is polarized by the MM charges. This sounds conceptually simple, but can be an intricate matter in practice. First of all, the QM density can become too delocalized due to interactions with the point charges, which is referred to as the "spill out problem", in particular when large basis sets of plane wave bases are used. This problem can be alleviated by using a modification of the $1/r$ interaction at short distances[33]. Further, large point charges close to the QM region can overpolarize the QM density due to the artificial concentration of the MM charge at one point. Here, a charge smearing scheme can be used[28]. Finally, in the *polarizable embedding* scheme a polarizable force field instead of the fixed point charge model is used. In some cases, polarization effects from the environment can have a significant impact on the result as shown, for example, by the calculation of excitation energies in retinal proteins[34,35] (see below).

3 Sampling Reactive Events

For chemical reactions, the calculation of free energy changes and activation free energies is of ultimate interest and is still a challenge. There are several categories of techniques available.

- **Direct MD** The most straightforward way is to perform MD simulations by integrating Newton's equation of motion with either the microcanonical or canonical ensembles.[36]. The common computational technology and algorithms, however, put severe limitations in the accessible time scales. As a rule of thumb, HF and DFT methods allow to perform MD simulations in the ps regime ($\approx$ 10-50ps for 'small' QM regions of 10-50 atoms), while SE methods allow for simulation times roughly three orders of magnitude longer ($\approx$ 10-100ns for 'small' QM regions). Therefore, direct MD simulations only allow overcoming small free energy barriers of several k_BT, such as sampling of various conformers of very short peptides in water (see below). Many chemical reactions of interest have barriers on the order of 10-25 kcal/mol, and can not be meaningfully addressed with direct MD simulations, even with SE methods. Direct MD simulations, therefore, are mostly useful for equilibrating configurations of protein active sites and qualitative exploration of the structural features relevant to chemistry, such as water distributions along potential proton transfer pathways.

- **Reaction path techniques** These methods determine the Minimum Energy Path (MEP)[37] between a reactant and product state, in particular they locate the transition state (saddle point on the potential energy surface). For enthalpy driven processes, this path contains most relevant information for describing the chemical reaction of interest, in particular the relative energies of reactant, product and transition state. As

a starting point, reactant and product states have to be available. For simple reactions, an approximate MEP can be determined by the adiabatic mapping procedure[38], when a reaction coordinate is chosen and partial optimizations are carried out with the reaction coordinate set to a number of values; e.g., consider a proton transfer from an oxygen atom to a nitrogen, denote the O-H distance by d_1, the H-N distance by d_2, a reaction coordinate $d = d_1 - d_2$ can then be used to describe the reaction. For more complex reaction processes that actively involve many degrees of freedom, however, more sophisticated techniques are required. One technique available in CHARMM is called the Conjugate Peak Refinement (CPR,[39]), which starts by a straight line interpolation between reactant and product. At the line search maximum, all degrees of freedom perpendicular ('conjugate') to the initial search direction are optimized, until a minimum is found. This minimum is then connected to the reactant and product and the optimization process is iterated. A popular alternative is the Nudged elastic band method (NEB[40]), where images of the system are distributed along the search line between reactant and product and are connected by springs; the related dimer method[41] is also widely used, though more in solid state and surface physics communities.

For enthalpy driven processes, MEP based techniques can provide valuable mechanistic information. The limitations of the methods, however, are also obvious. First, the straight line interpolation does not assure to find the pathway with lowest energy [a]. Therefore, chemical intuition is necessary to include various different intermediate states, as illustrated in our study of the first proton transfer event in Bacteriorhodopsin[42] (see below). Moreover, entropic contributions are completely neglected. For example, Klähn et al.[43] showed for the reaction of a phosphate ion in the Ras-GAP complex that the total energies of reactant and product fluctuate on the order of 30 kcal/mole and the reaction barrier on the order of 6 kcal/mol, when using different protein conformations generated by classical MD simulations. In other words, the thermal motion of the protein environment makes the use of total energies in the MEP framework meaningless, which highlights the point that pursuing a high accuracy in the QM method may not be the bottleneck for meaningful QM/MM studies of many biological problems.

- **Free energy computations along reaction path** One approach for improving upon MEP results is to calculate the free energy (potential of mean force) along the MEP. For example, the MM free energy contribution along the MEP can be estimated using free energy perturbation calculations in which the QM region is frozen (or treated in a harmonic fashion) while the MM region samples the proper thermal distribution orthogonal to the MEP[44]. In the more elaborate scheme developed recently[45], the path itself can be refined based on the derivatives of the potential of mean force, which ultimately converges to a minimum free energy path. The cost of such calculations, however, can be rather high especially if high-level QM methods are used; one practical approximation is to replace the QM region by effective (or even polarizable) charges when sampling the MM degrees of freedom[46].

- **Umbrella sampling and meta-dynamics** When the reaction can be described by a

[a] Imagine connecting Munich and Milano by a rope, which will arrange along the valleys connecting Munich and Milano: however, depending on the initial placement of the rope, different pathways can be found.

number of pre-chosen "reaction coordinates", umbrella sampling techniques[47] can be used to generate the relevant potential of mean force curve/surface. The most basic technique is to add harmonic umbrella potentials at a discrete set of reaction coordinate values to overcome barriers on the potential energy surface, and various schemes have been proposed to make the process automated (adaptive) and converge quickly. For example, meta-dynamics methods[48] are adaptive umbrella sampling methods where successive Gaussians are added to avoid revisiting configurations during the sampling and therefore speeds up the convergence; the width and height of the added Gaussian functions as well as the frequency of adding the Gaussian functions can be optimized for optimal convergence[49–51]. Finally, energy can be used as a collective reaction coordinate to enhance sampling when it is difficult to determine *a priori* a set of geometrical parameters that describe the reaction[52–54].

- **Other advanced techniques** Finally, there are transition path sampling (TPS) techniques that aim to directly sample the reactive trajectory ensembles[55]. These are in principle the most rigorous framework for understanding reaction mechanisms in the condensed phase and generally do not require specifying *a priori* the reaction coordinates; it is well known that environmental degrees of freedom can be essential part of the kinetic bottleneck for many reactions in solution and biological systems. TPS has been applied in several studies of enzyme reactions[56,57], and the cost of such calculations highlights the importance of developing accurate SE methods. It should be noted that the TPS techniques in principle can also suffer from sampling issues in the path space and therefore can also benefit from using different initial guesses.

4 Semi-Empirical Methods

While the adiabatic mapping calculations can be readily applied in conjunction with HF and DFT methods, more elaborate reaction path techniques and most free energy and TPS techniques overstretch the possibilities of *ab initio* methods and are mostly applied using SE methods. The great promise of DFT methods on the one hand and the lower accuracy and limited transferability of the SE type methods, like MNDO, AM1 or PM3 on the other hand, seemed to devalue the latter type of methods; in the late 1990's they were to become obsolete in the eyes of many quantum chemist's. However, the limitations and quite involved empirical parametrization process of modern DFT methods changed also the view onto the SE methods[58]. The desire to study increasingly complex (bio)molecules and the importance of entropic contribution and sampling in studying soft matter brought a renewed interest into SE methods, especially if they can be made more robust and transferable.

Most SE methods are derived from the Hartree-Fock theory by applying various approximations resulting in, for example, the Neglect of Differential Diatomic Overlap (NDDO) type of methods; the most well-known ones being the MNDO, AM1 and PM3 models[59]. In these methods certain integrals are omitted and the remaining are treated as parameters, which are either pre-calculated from first principles or fitted to experimental data. SE methods usually have an overall accuracy lower than DFT, although this can be reversed for specific systems. In the so called specific reaction parametrization (SRP) scheme[60], a SE method (e.g., PM3) is specifically re-parametrized for the particular sys-

tem under study, which may provide a very accurate description for the reaction of interest at a level even unmatched by popular DFT methods. However, parameterization of a SRP that works well for condensed phase simulations is not as straightforward as for gas-phase applications and a large number of carefully constructed benchmark calculations are needed[61–63]. Therefore, it remains an interesting challenge to develop generally robust SE methods that properly balance computational efficiency and accuracy. Some of the more recent models include the inclusion of orthogonalization corrections in the OMx model[59], the PDDG/PM3 model[64], and the NO-MNDO model, which all generated encouraging improvements over traditional NDDO methods.

SE methods can also be derived from DFT, a development that we have focussed on over the last decade. The so called Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) method[65,66] is derived by expanding the DFT total energy functional up to second order with respect to the charge density fluctuations $\delta\rho$ around the reference density ρ_0[66] ($\rho_0' = \rho_0(\vec{r}')$, $\int' = \int d\vec{r}'$):

$$E = \sum_i^{occ} \langle \Phi_i | \hat{H}^0 | \Phi_i \rangle + \frac{1}{2} \iint' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \left. \frac{\delta^2 E_{xc}}{\delta\rho\, \delta\rho'} \right|_{\rho_0} \right) \delta\rho\, \delta\rho' - \frac{1}{2} \iint' \frac{\rho_0' \rho_0}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0]\rho_0 + E_{cc} \tag{3}$$

$\hat{H}^0 = \hat{H}[\rho_0]$ is the effective Kohn-Sham Hamiltonian evaluated at the reference density ρ_0 and the Φ_i are Kohn-Sham orbitals. E_{xc} and V_{xc} are the exchange-correlation energy and potential, respectively, and E_{cc} is the core-core repulsion energy (an extension up to third order has been presented recently[67,68]).

The (artificial) reference density ρ_0 is chosen as a superposition of densities ρ_0^α of the neutral atoms α constituting the molecular system,

$$\rho_0 = \sum_\alpha \rho_0^\alpha \tag{4}$$

and a density fluctuation $\delta\rho$, also built up from atomic contributions

$$\delta\rho = \sum_\alpha \delta\rho^\alpha, \tag{5}$$

in order to represent the ground state density

$$\rho = \rho_0 + \delta\rho. \tag{6}$$

Approximations to the three energy contributions in eq. 3 result in the final expression of the SCC-DFTB model[66]:

$$E = \sum_{i\mu\nu}^{occ} c_\mu^i c_\nu^i < \eta_\mu | \hat{H}^0 | \eta_\nu > + \frac{1}{2} \sum_{\alpha,\beta} U_{\alpha\beta}(R_{\alpha\beta}) + \frac{1}{2} \sum_{\alpha\beta} \Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta} \tag{7}$$

SCC-DFTB has been tested in detail for atomization energies, geometries and vibrational frequencies using a large set of molecules[69–71]. In terms of atomization energies, the modern NDDO type methods like PDDG/PM2 or OM2 have bee shown to be superior to SCC-DFTB, while SCC-DFTB is excellent in reproducing geometries and also predicts reasonable vibrational frequencies. It is worth emphasizing again that the SE methods are likely less accurate than modern DFT-functionals on average, although this situation can be reversed in specific cases.[b] Moreover, as discussed above, the errors introduce by neglecting the effects of dynamics and entropy can become larger than the intrinsic error of the respective electronic structure method. Nanoseconds of MD simulations are readily feasible with SE methods, while impossible with HF and DFT. Therefore, SE methods can be used in various ways to improve the quality of the computational model: (i) They can be applied as the main QM method for the initial exploration of possible reaction mechanisms after careful testing/refinement for relevant model systems; (ii) they can be used to estimate the entropic contributions of a particular mechanism while the accurate potential energy is evaluated at a higher level method[77]; (iii) they can be used as the lower level QM in either an ONIOM type multi-level[23,24] scheme or to guide the sampling in a multi-level free energy calculations.

5 The Continuum Component

While continuum approaches applied in computational materials science mostly model mechanical properties, those applied in biological simulations mainly model the dielectric response of the environment to the charge distribution of the molecule[c]. Most popular continuum electrostatics models in the biological context are based on the Poisson-Boltzmann framework.[11,13]. The Poisson equation allows to compute the electrostatic potential and electrostatic free energy for the charge distribution of the solute in the presence of a dielectric continuum (representing the dielectric response from the solvent molecules). The PB equation further includes the mobile charge distribution originating from the surrounding ions, which respond to and modulate the electrostatic potentials of the solute charges:

$$-\nabla\epsilon(x)\nabla\phi(x) - \sum_{k=1}^{N} c_k q_k e^{-q_k\phi(x)-V_k(x)} = \frac{4\pi e^2}{kT}\rho(x) \tag{8}$$

$\rho(x)$ describes the solute charge distribution, q_k the charge of the mobile ion species k, $V_k(x)$ the steric interaction of the solute with the mobile ion species k, $\epsilon(x)$ the space-dependent dielectric function and $\phi(x)$ the resulting electrostatic potential. As discussed extensively in the literature[13], the correlation between the mobile ions is ignored in the PB approach, thus PB is most reliable for monovalent ions, which fortunately fits most biological applications.

[b] Even well established methods like the hybrid DFT method B3LYP show deficiencies, which may not be widely recognized, e.g., problems with the description of extended electronic π systems[72,73], dispersion interactions[74] or electronically excited states with significant charge-separation[75,76]. These examples show that careful testing are obligatory before application to new systems, even for DFT methods.

[c] Generally, the work for cavity formation and the van der Waals interactions at the surface, the 'apolar' components of the solute-solvent interaction, need to be included as well, see[78].

In the most straightforward conceptual scheme, the QM/MM/CM treats the active site with QM, the entire biomolecule with MM and the solvent with CM. In many practical applications, however, it is sufficient to treat atoms very close to the active site (e.g., within 20 Å) with discrete MM degree of freedom that are fully flexible during the simulation; this would include explicit solvent molecules in or near the active site, which helps alleviate some of the limitations of continuum electrostatics models at the solute/solvent interface. To properly and efficiently deal with protein atoms in the continuum region, we have adapted the Generalized Solvent Boundary Condition (GSBP) scheme developed by Roux and co-workers for classical simulations[79]. Briefly, if we refer the discrete QM/MM region as the inner region while the continuum as the outer regions, the total effective potential (potential of mean force) of the system can be written as,

$$W_{GSBP} = U^{(ii)} + U^{(io)}_{int} + U^{(io)}_{LJ} + \Delta W_{np} + \Delta W^{(io)}_{elec} + \Delta W^{(ii)}_{elec}, \tag{9}$$

where $U^{(ii)}$ is the complete inner-inner potential energy, $U^{(io)}_{int}$ and $U^{(io)}_{LJ}$ are the inner-outer internal (bonds, angles, and dihedrals) and Lennard-Jones potential energies, respectively, and ΔW_{np} is the non-polar confining potential. The last two terms in Eq.9 are the core of GSBP, representing the long-range electrostatic interaction between the outer and inner regions. The contribution from distant protein charges (screened by the bulk solvent) in the outer region, $\Delta W^{(io)}_{elec}$, is represented in terms of the corresponding electrostatic potential in the inner region, $\phi^{(o)}_s(\mathbf{r}_\alpha)$,

$$\Delta W^{(io)}_{elec} = \sum_{\alpha \in inner} q_\alpha \phi^{(o)}_s(\mathbf{r}_\alpha) \tag{10}$$

The dielectric effect on the interactions among inner region atoms is represented through a reaction field term,

$$\Delta W^{(ii)}_{elec} = \frac{1}{2} \sum_{mn} Q_m M_{mn} Q_n \tag{11}$$

where $\mathbf{M}$ and Q are the generalized reaction field matrix and generalized multipole moments, respectively, in a basis set expansion.[79]

The advantage of the GSBP method lies in its ability to include these contributions explicitly while sampling configurational space of the reaction region during a simulation at minimal additional cost. The static field potential, $\phi^{(o)}_s(\mathbf{r})$, and the generalized reaction field matrix $\mathbf{M}$ are computed only once based on PB calculations and stored for subsequent simulations. The only quantities that need to be updated during the simulation are the generalized multipole moments, Q_n,

$$Q_n = \sum_{\alpha \in inner} q_\alpha b_n(\mathbf{r}_\alpha) \tag{12}$$

where $b_n(\mathbf{r}_\alpha)$ is the nth basis function at nuclear position $\mathbf{r}_\alpha$.

As described in Ref.[18], the implementation of GSBP into a combined QM/MM framework is straightforward, and involves the QM-QM and QM-MM reaction field, and the QM-static field terms. For the GSBP combined with SCC-DFTB, these terms take on a simple form because $\rho^{QM}(\mathbf{r})$ is expressed in terms of Mulliken charges.[66] Although the formulation of GSBP is self-consistent, the validity of the approach depends on many factors especially the size of the inner region and the choice of the dielectric "constant" for

the outer region. Therefore, for any specific application, the simulation protocol has to be carefully tested using relevant benchmarks such as pK_a of key residues[17,80].

An economic alternative to the GSBP approach is the charge-scaling protocol, where the partial charges for charged-residues in the MM region are scaled down based on the electrostatic potentials calculated (with PB) when the biomolecule is in vacuum vs. solution; the scaled partial charges are then used in the QM/MM simulations with the system in vacuum. In the end, PB calculations are carried out with scaled and full partial charges to complete the thermodynamic cycle. The charge-scaling approach has been successfully used in several QM/MM studies[7,81,82], although several numerical issues (e.g., treatment of residues very close to the QM region and cancellation of large contributions) render the protocol less robust than GSBP.

6 Polarizable Force Field Models

Continuum electrostatic approaches take into account a majority of the dielectric responses of the solute. The electronic polarization of the environment to changes in the QM density during chemical reaction, however, is missing when non-polarizable force fields are used as MM. This electronic polarization may give significant contributions when electrons or ions are transported over long distances and for excitation energies, where the dipole moment of the QM region changes significantly upon excitation. In the last decade, many research groups have been actively developing polarizable force fields, for which several good overviews are available (see a thematic issue that follows *J. Chem. Theory Comput.* 2007, 3, 1877).

Common approaches to describe electronic polarization effects use models based on atomic polarizabilities, a method that we have implemented to estimate the protein polarization effects on electronic excitation energies[35]. Here, the Coulomb interaction is described using atomic charges q_A and atomic polarizabilities α_A, where the induced atomic dipoles μ_A can be calculated as:

$$\mu_A = \alpha_A \left(\sum_B \mathbf{T_{AB}} q_B + \sum_C \mathbf{T_{AC}} \mu_C \right) \tag{13}$$

The first term contains the Coulomb interaction with the fixed atomic point charges, which lead to the induced dipoles. The second term describes the interaction between the induced dipoles. Note that the induced dipole moments appear on both sides of the equation. For small systems, these equations can be solved by matrix inversion techniques, for large systems they are usually solved iteratively. The tensors $\mathbf{T}$ contain a damped Coulomb interaction since for small distances the bare Coulomb $1/r$ and $1/r^3$ terms for the charge-charge and charge-dipole interactions would lead to over-polarization. Effectively, this damping is induced by smearing out the charges, i.e., by describing the atomic charge with an exponential charge distribution. A new parameter, the width 'a' of this charge distribution is therefore introduced and has to be determined during the fitting procedure.

Atomic polarizabilities can be calculated or taken from experiment; we have used values from the literature[35], where typical parameters are around 0.5 $\mathring{A}^{-3}$ for H and about 0.8-1.5 $\mathring{A}^{-3}$ for first row atoms C, N and O. However, to gain high accuracy atomic parameters have been taken to be dependent on the atomic hybridization state, e.g., the parameters for

sp^2 and sp^3 carbon differ by about 0.5 $\mathring{A}^{-3}$. This allows to account for the different polarizabilities of sp^3 carbon structures, like alkanes compared to aromatic molecules, like polyenes or benzene.

Common force charge models are parametrized in order to account for the effects of solvation implicitly. This can be done by fitting the charges to experimental data, or by calculating them using HF/6-31G*, which is known to overestimate the magnitude of charges, thereby implicitly taking the effect of solvent polarization into account. Therefore, as a first step a new charge model has to be developed in order to be consistent with an explicit treatment of polarization. We computed 'polarization free' charges by performing B3LYP/6-311G(2d,2p) calculations and fitting the charges to produce the electrostatic potential at certain points at the molecular boundary (RESP)[35]. [d] These 'polarization free charges' are computed for certain molecular fragments in gas phase, i.e. for certain chemical groups like amino acid side chains. The charges therefore already contain the mutual polarization within these fragments. Therefore, the polarization model is also restricted interactions between these fragments and not applied within one region to avoid double counting.

Critical tests include the calculation of polarizability tensors of amino acid side chains in comparison with DFT and MP2 data, and the evaluation of the polarization energy of such side chains due to a probe charge in the vicinity. The polarization model is able to reproduce the QM data with high precision[35], allowing therefore for meaningful calculations on larger systems like entire proteins.

7 Applications

In this section, we discuss three applications, to illustrate the various methodologies discussed above.

7.1 Direct QM/MM MD with periodic boundary conditions: Dynamics of peptides and proteins

The conformation of peptides and proteins depend sensitively on the proper inclusion of solvent. The conformations of small peptides in the gas phase are very different from those in solution and it is challenging to use a QM description of the peptide augmented with an implicit solvent model to model those properly. One possible approach is to include the first solvation shell explicitly[83], although finite temperature effects still need to be included, which can be problematic with a small "microsolvation" model. A physically more transparent model is to surround the peptide, treated with QM, by a box of MM water molecules and to apply periodic boundary conditions[84]. The main degrees of freedom in these peptides are the backbone torsions (ϕ, ψ) , which exhibit rotational barriers of a few kcal/mol (Fig.1). To sample the energy landscape of such systems, MD simulations in the order of 10-100 nano-seconds have to be performed, which is clearly only possible using SE methods. This also illustrates the limits of direct MD simulations, which can handle only systems with small barriers of a few kcal/mol. Linear scaling methods in combination with SE methods allow to simulate the dynamics of small proteins over several 100 ps[85].

[d] Diffuse basis functions should be avoided, since those would allow the charge density into regions far away from the molecule, which are not accessible in the condensed phase due to the environment.

However, this is still quite costly and there are not too many applications where a QM treatment of the entire protein is necessary and the dynamics on these short time-scales are the quantities of key interest.

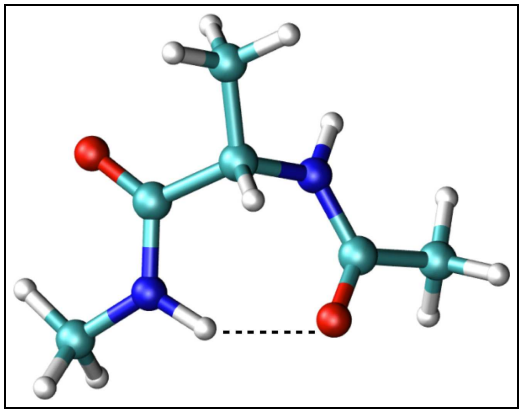

Figure 1. The lowest energy conformation C_7^{eq} of the alanine dipetide model in the gas phase. The main degrees of freedom consist of the phi and psi dihedral angles, i.e., rotations around the central C-C and C-N single bonds.

7.2 Proton transfer

Proton transfer reactions are involved in many key biological problems, most notably in acid-base catalysis and bioenergetics processes. The breaking and formation of many chemical bonds in these problems and the significant reorganization of the environment in response to the transport of charges pose great challenges to theoretical studies. Although more specialized techniques such as MS-EVB can be extremely valuable in the study of certain proton transfer problems[86], a QM/MM framework is required to introduce more flexibility in the potential energy surface, especially when the reaction involves species of complex electronic structures (e.g., transition metal ion). The diversity of proton transfer reactions also makes them ideal for illustrating the value and limitation of various QM/MM techniques.

7.2.1 Bacteriorhodopsin (bR): MEP results

For relatively localized proton transfers, for which the entropic contribution is likely small, reaction path methods can be applied. An example is the first proton transfer step in bacteriorhodopsin, where the active site involves well connected hydrogen bonding network as shown in Fig.2. It is known from experiment that entropy does not contribute to this step, therefore, we have simulated the process using SCC-DFTB QM/MM in combination with the CPR approach discussed above[42,87]. The computed barriers of 11.5-13.6 kcal/mol for different low-energy pathways are in good agreement with the experimental value of 13 kcal/mol. However, to understand this properly one has to be aware of the intrinsic error compensation in these calculations: as discussed in detail in Ref.[88], popular DFT methods

tend to underestimate proton transfer barriers by 1-4 kcal/mole. On the other hand, the inclusion of nuclear quantum effects like zero point energies would lower proton transfer barriers by roughly this amount, therefore, these two effects tend to cancel each other for a wide range of proton transfer systems.

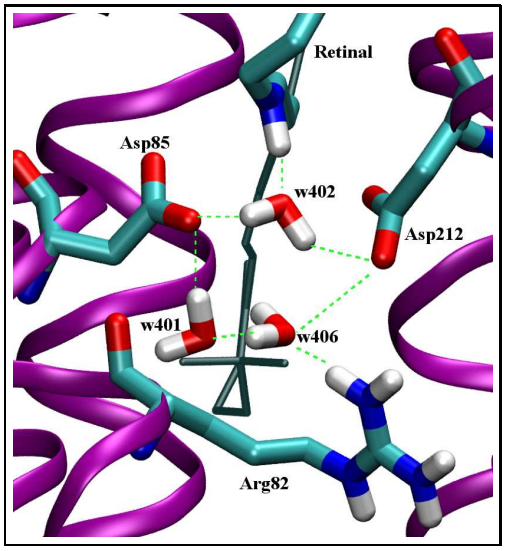

Figure 2. The active site of bacteriorhodopsin in its ground state. The first proton transfer occurs between the retinal Schiff base and the side chain Asp85.

7.2.2 Carbonic Anhydrase II : MEP vs. PMF

For many long-range proton transfers in biomolecules, however, the MEP results are likely very sensitive to the protein structure used in the calculation. More severely, the collective structural response in the protein is likely missing in the MEP calculations, which may lead to qualitatively incorrect mechanistic conclusions. A useful example in this context is the long-range proton transfer in carbonic anhydrase II (CAII), where the rate-limiting step of the catalytic cycle is a proton transfer between a zinc-bound water/hydroxide and the neutral/protonated His64 residue close to the protein/solvent interface. Since this proton transfer spans at least 8-10 Å, the transfer is believed to be mediated by the water molecules in the active site[89] (see Fig.3). Since there are multiple water wires of different length in the active site that connect the donor/acceptor groups (zinc-bound water, His 64), a question of interest is whether specific length of water wire dominates the proton transfer or all wires have comparable contributions.

First, a large number of MEPs have been collected starting from different snapshots collected from equilibrium MD simulations at the SCC-DFTB/MM level. Since essentially a positive charge is transferred over a long-distance, it was expected that the MEP energetics depend sensitively on the starting structure, which was indeed observed. For example,

when the starting structure came from a CHOH (zinc-bound water, neutral His64) trajectory, the proton transfer from the zinc-water to His64 is largely *endothermic* (on average by as much as $\sim$ 13 kcal/mol). By contrast, when the starting structure came from a COHH (zinc-bound hydroxide, protonated His64) trajectory, the same proton transfer reaction was found largely *exothermic*. As an attempt to capture the "intrinsic barrier" for the proton transfer reaction, which is known to be close to be thermoneutral experimentally,[90] we generated configurations from equilibrium MD simulations in which protons along a specific type of water wire were restrained to be equal distance from nearby heavy atoms (e.g., oxygen in water or $N\epsilon$ in His 64). In this way, the charge distribution associated with the reactive components is midway between the CHOH and COHH states, thus the active-site configuration was expected to facilitate a thermoneutral proton transfer process, which was indeed confirmed by MEP calculations using such generated configurations as the starting structure. An interesting observation is that the barriers in such "TS-reorganized" MEPs showed a steep dependence on the length of the water wire; it was small ($\sim$ 6.8$\pm$2.2 kcal/-mol) with short wires but substantially higher than the experimental value ($\sim$ 10 kcal/mol) with longer water wires (e.g., 17.4$\pm$2.0 kcal/mol for four-water wires).

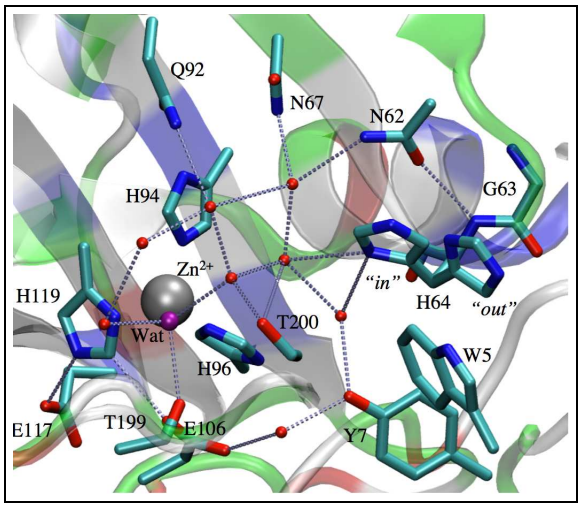

Figure 3. The active site of CAII rendered from the crystal structure (PDB ID: 2CBA[89]). All dotted lines correspond to hydrogen-bonding interactions with distances $\leq$3.5 Å. The proton acceptor, His64, is resolved to partially occupy both the "in" and "out" rotameric states.

This steep wire-length dependence is in striking contrast with the more rigorous PMF calculations.[91,92] In the PMF calculations, a collective coordinate[93] was used to monitor the progress of the proton transfer without enforcing specific sequence of events involving individual protons along the wire; the use of a collective coordinate is important because this allows averaging over different water wire configurations, which is proper since the life-time of various water wires is on the pico-second time scale,[18,20] much faster than the time scale of the proton transfer (μs).[90] In the PMF calculations, the wire-length dependence was examined by comparing results with different His 64 orientations ("in" and

"out", which is about 8 and 11 Å from the zinc, respectively); both configurations were found to involve multiple lengths of water wires but different relative populations. The two sets of PMF calculations produced barriers of very similar values, which suggested that the length of the water wire (or orientation of the acceptor group) is unlikely essential to the proton transfer rate. Further analysis of the configurations sampled in the MEP simulations suggested that the MEP results artificially favored the concerted proton transfers, which correlate to significant distance dependence. As discussed above, to generate the "TS-reorganized" configurations, all transferring protons along the wire were constrained to be half-way between the neighboring heavy atoms; therefore, such sampled protein/-solvent configurations would favor a concerted over step-wise proton transfers. Although *all* atoms in the inner region are allowed to move in the MEP searches, the local nature of MEPs does not allow collective reorganization of the active site residues/solvent molecules thus the "memory" of the sampling procedure is not erased.

Therefore, the CAII example clearly illustrates that care must be exercised when using MEP to probe the mechanism of chemical reactions in biomolecules, especially when collective rearrangements in the environment are expected (e.g., reactions involving charge transport). Along the same line, the GSBP based QM/MM/CM framework was found to be particularly attractive in the CAII studies for maintaining the proper solvent configurations and sidechain orientations in the active site, as compared to Ewald based SCC-DFTB/MM simulations[18,80], at a faction of the computational cost. Ignoring the bulk solvation effect, for example, was found to lead to unphysical orientations of the functionally important His64 residue.

7.2.3 Membrane proteins

A particularly exciting area for which the multiscale QM/MM/CM approach is suited concerns proton translocation across membrane proteins, where a proper and efficient treatment of the heterogeneous protein/solvent/membrane environment is particularly important, such as in bacteriorhodopsin and cytochrome c oxidase. The GSBP framework also allows one to incorporate the effect of membrane potential[94], which plays a major role in bioenergetics, in a numerically efficient manner. Using the SCC-DFTB/MM/GSBP protocol with a relatively small inner region ($\sim$ 30Å$\times$30Å$\times$50Å) and dielectric membrane model[93], we were able to reproduce the water wire configurations in the interior of aquaporin in good agreement with the much more elaborate MD simulations using four copies of aquaporin embedded in an explicit lipid bilayer. Ignoring the GSBP contributions, however, led to very different water distributions, which highlights the importance and reliability of the multiscale framework. In a recent study[95], the same framework was also found semi-quantitatively successful in predicting pK_a of titritable groups in the interior of bacteriorhodopsin and cytochrome c oxidase, which are extremely challenging and relevant benchmark for studying proton transfer systems in general[96]. Finally, the SCC-DFTB/MM/GSBP studies of the proton release group (PRG) in bacteriorhodopsin[97] led to the key insight that the PRG is not a protonated water cluster as proposed in a series of recent IR studies[98,99]; rather, the PRG is a pair of conserved Glutamate bonded together with a delocalized proton (see Fig.4), and it is the delocalization of this "intermolecular proton bond" that leads to the unusual IR signature found in experiments[98,99].

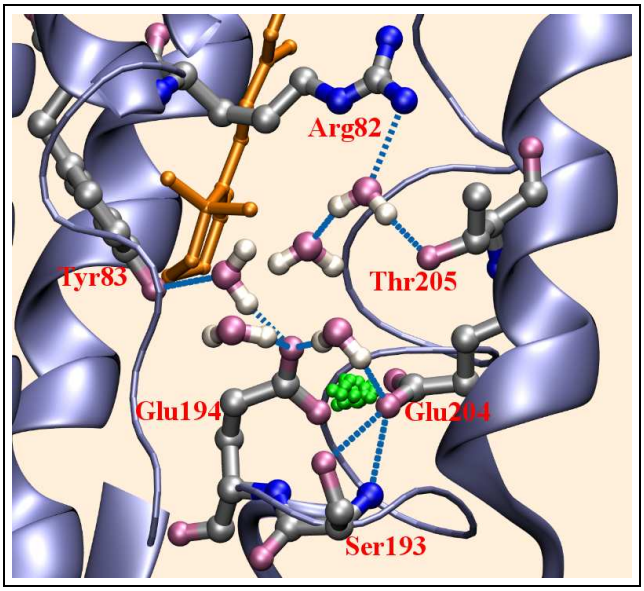

Figure 4. SCC-DFTB/MM-GSBP simulations indicate that the stored proton in the proton pump bacteriorhodopsin is delocalized (green spheres) between a pair of conserved glutamate residues rather than among the active site water molecules.

7.3 Excited states properties

The accurate determination of excited states properties is a challenging task for quantum chemical methods in general. This holds true in particular for the chromophore in retinal proteins (like bR), a polyene chain linked via a Schiff-base (NH) group to the protein backbone[76, 73, 100] (see Fig.5). Due to its extended and highly-correlatedπ-electron system, retinal is highly polarizable and undergoes a large change in dipole moment upon excitation, therefore, protein polarization effects may become important for an accurate description of excited state properties.

Standard QM/MM calculations using only an electrostatic embedding scheme do not take the (electronic) polarization response of the protein environment into account, which is different for ground and excited states due to the change of the dipole moment upon excitation. [e] In the case of retinal, the dipole in the excited state is about 10 Debye larger than in the ground state, therefore, MM polarization stabilizes the excited state more than the ground state, leading to an effective red-shift in excitation energies.

Indeed, QM/MM electrostatic embedding calculations tend to overestimate the excitation energy. While the experimental absorption maximum is at 2.18 eV, MRCI QM/MM calculations estimate it to be 2.34 eV, other methods predict even more blue shifted values[73]. There are many factors that contribute to the computational uncertainty, one of which being the intrinsic accuracy of the applied QM method. Other factors are related to the QM/MM coupling and the electrostatic treatment of the environment. For example,

[e]They of course can take the 'ionic' response into account, i.e., the relaxation of the protein structure, which also leads to a change in the electrostatic field from the MM environment.

different force field models (like AMBER and CHARMM) use different point charge models, which can lead to differences in the excitation energies on the order of 0.05 eV[35]. In many applications, only the protein is included in the MM treatment, the larger environment including membrane and bulk water is neglected. This effect can be estimated with a linearized version of the Poisson Boltzmann equation 8 in the charge scaling[81] approach as discussed above. Estimating excitation energies with and without charges scaling results again in differences of about 0.05 eV.

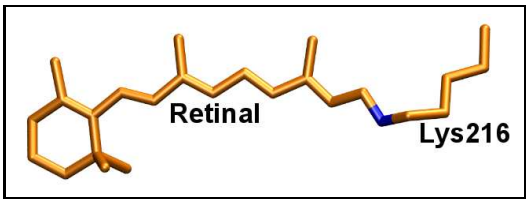

Figure 5. The retinal chromophore in the all-$trans$ conformation, as in the bR ground state. The blue color indicates the Schiff base (NH) group, from which the proton is transferred in the first step to Asp85.

Using a polarizable model, the ground state charge distribution in the MM region is determined using eq. 13. The resulting charges may be different from those in the regular force field models, because they are computed in response to the actual electrostatic field of the protein with retinal in the ground state. This charge distribution leads to vertical excitation energies about 0.07 eV red-shifted compared to those from the CHARMM force field[35]. In the same way, a different set of MM charges can be determined for the case where retinal is in its excited state. This change in the electrostatic environment leads to a further red shit of 0.07 eV, which is due to the different MM polarization in the ground and excited states. The total red-shift with respect to the CHARMM charges is 0.14eV, showing that protein polarization can have a significant impact on excitation energies in those cases, where the dipole moment of the chromophore changes significantly upon excitation.

A different approach to estimate the effect of polarization is to use a low level QM method instead of the polarizable MM region. We have used such a QM/QM'/MM approach, applying charge scaling, a MRCI method for the QM region containing the retinal chromophore and a DFT methods for 300 atoms around the chromophore in the QM' region to benchmark the polarizable MM model[34]. This study showed that the well-calibrated polarizable MM model gives nearly the same results as the QM' region. However, the 300 atom QM' region leads only to roughly 50% of the red-shift, showing that a large MM region contributes to the polarization effect.

8 Summary

In the last decade, many variants of multiscale methods have been developed to study chemical events in complex environments in materials science, chemistry and biology. The specific design of such methods depends very much on the properties of the investigated system and the problem in hand. Biological systems are characterized by their high degree of structural flexibility and the long-range nature of the electrostatic forces, which

are essential to the understanding of biological functions. Therefore, the main emphasis in methods development in the biological context lies in the accurate representation of electrostatics and algorithms to tackle the sampling problem. In this article, we have discussed QM/MM algorithms embedded into an implicit electrostatic environment, which is modeled based on the Poisson-Boltzmann equation. For many applications, the representation of the MM environment by fixed point charges may be appropriate, however, in cases where the electrostatic properties in the QM region change significantly, a polarizable MM representation is likely required. Thermal fluctuations, on the other hand, can lead to a significant contribution to the free energies that characterize the chemical reaction. Accordingly, expensive QM methods often have to be substituted by more efficient, although less accurate ones. We have described applications using various approximations for the QM region. For the determination of excitation energies, high level QM methods have to be applied, while for the study of proton transfer events, DFT and approximate SCC-DFTB can lead to a balanced treatment allowing to draw meaningful conclusions about the reaction mechanism and energetics. In some cases, the neglect of thermal fluctuations would even lead to much larger errors than the use of lower accuracy QM methods. Therefore, studying biological systems requires applying a multitude of methods and calculating multiple experimental observables to reach reliable mechanistic conclusions.

Acknowledgments

We are indebted to our collaborators for their contributions, without which the work described here can't be accomplished. Supports from the National Science Foundation, National Institutes of Health, Alfred P. Sloan Foundation, DFG and computational resources from the National Center for Supercomputing Applications at the University of Illinois are greatly appreciated.

References

1. M. Karplus and J. A. McCammon, *Molecular dynamics simulations of biomolecules*, Nat. Struct. Mol. Biol., **9**, 646–652, 2002.
2. W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glattli, P. H. Hünenberger, M. A. Kastenholz, C. Ostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, and H. B. Yu, *Biomolecular modeling: Goals, problems, perspectives*, Angew. Chem. Int. Ed., **45**, 4064–4092, 2006.
3. A. D. MacKerell Jr., D. Bashford, M. Bellot, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III., B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins*, J. Phys. Chem. B, **102**, 3586–3616, 1998.
4. W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, *Biomolecular Simulation: The GROMOS Manual and User Guide.*, vdf Hochschulverlag, ETH Zürich, Switzerland, 1996.

5. J. W. Ponder and D. A. Case, *Force fields for protein simulations*, Adv. Prot. Chem., **66**, 27, 2003.
6. A. Warshel, *Computer simulations of enzyme catalysis: Methods, Progress and Insights*, Annu. Rev. Biophys. Biomol. Struct., **32**, 425–443, 2003.
7. Qiang Cui and Martin Karplus, *Catalysis and specificity in enzymes: A study of triosephosphate isomerase (TIM) and comparison with methylglyoxal synthase (MGS)*, Adv. Prot. Chem., **66**, 315–372, 2003.
8. J. L. Gao, S. H. Ma, D. T. Major, K. Nam, J. Z. Pu, and D. G. Truhlar, *Mechanisms and free energies of enzymatic reactions*, Chem. Rev., **106**, 3188–3209, 2006.
9. M. Cossi, V. Barone, R. Cammi, and J. Tomasi, *Ab initio study of solvated molecules: A new implementation of the polarizable continuum model*, Chem. Phys. Lett., **255**, 327–335, 1996.
10. C. J. Cramer and D. G. Truhlar, *Implicit solvation models: Equilibria, structure, spectra, and dynamics*, Chem. Rev., **99**, 2161–2200, 1999.
11. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, *Electrostatics of nanosystems: Application to microtubules and the ribosome*, Proc. Acad. Natl. Sci. USA, **98**, 10037–10041, 2001.
12. M. Feig and C. L. Brooks, *Recent advances in the development and application of implicit solvent models in biomolecule simulations*, Curr. Opin. Struct. Biol., **14**, 217–224, 2004.
13. J. P. Hansen and I. R. McDonald, *Theory of simple liquids, 3rd Ed.*, Academic Press, London, UK, 2006.
14. D. Bashford and D. A. Case, *Generalized born models of macromolecular solvation effects*, Annu. Rev. Phys. Chem., **51**, 129–152, 2000.
15. Warshel, A. and Levitt, M., *Theoretical Studies of Enzymic Reactions*, J. Mol. Biol., 1976.
16. K. Nam, J. Gao, and D. M. York, *An efficient linear-scaling ewald method for long-range electrostatic interactions in combined QM/MM calculations*, J. Chem. Theo. Comp., **1**, 2–13, 2005.
17. D. Riccardi, P. Schaefer, and Q. Cui, *pK_a calculations in solution and proteins with QM/MM free energy perturbation simulations: A quantitative test of QM/MM protocols*, J. Phys. Chem. B, **109**, 17715–17733, 2005.
18. P. Schaefer, D. Riccardi, and Q. Cui, *Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules*, J. Chem. Phys., **123**, 014905, 2005.
19. B. A. Gregersen and D. M. York, *Variational Electrostatic projection (VEP) methods for efficient modeling of the macromolecular electrostatic and solvation environment in activated dynamics simulations*, J. Phys. Chem. B, **109**, 536–556, 2005.
20. D. Riccardi, P. Schaefer, Y. Yang, H. Yu, N. Ghosh, X. Prat-Resina, Peter Konig, G. Li, D. Xu, H. Guo, M. Elstner, and Qiang Cui, *Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes (Feature Article)*, J. Phys. Chem. B, **110**, 6458–6469, 2006.
21. H. M. Senn and W. Thiel, *QM/MM studies of enzymes*, Curr. Opin. Chem. Biol., **11**, 182–187, 2007.
22. F. Maseras and K. Morokuma, *IMOMM - A new integrated ab initio plus molecular mechanics geometry optimization scheme of equilibrium structures and transition states*, J. Comp. Chem., **16**, 1170–1179, 1995.

23. M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, *ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition*, J. Phys. Chem., **100**, 19357–19363, 1996.
24. Q. Cui, H. Guo, and M. Karplus, *Combining ab initio and density functional theories with semiempirical methods*, J. Chem. Phys., **117**, 5617–5631, 2002.
25. M. J. Field, P. A. Bash, and M. Karplus, *A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations*, J. Comput. Chem., **11**, 700–733, 1990.
26. Y. Zhang, T.S. Lee, and W. Yang, *A pseudobond approach to combining quantum mechanical and molecular mechanical methods*, J. Chem. Phys., **110**, 46–54, 1999.
27. J. Gao, P. Amara, C. Alhambra, and M. J. Field, *A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations*, J. Phys. Chem. A, **102**, 4714–4721, 1998.
28. D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chattfield, M. Hodošček, and B. R. Brooks, *Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method*, J. Chem. Phys., **117**, 10534–10547, 2002.
29. N. Reuter, A. Dejaegere, B. Maigret, and M. Karplus, *Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches*, J. Phys. Chem. A, **104**, 1720–1735, 2000.
30. I. Antes and W. Thiel, *Adjusted Connection Atoms for Combined Quantum Mechanical and Molecular Mechanical Methods.*, J. Phys. Chem. A, **103**, 9290, 1999.
31. P. H. König, M. Hoffmann, Th. Frauenheim, and Q. Cui, *A critical evaluation of different QM/MM frontier treatments using SCC-DFTB as the QM method*, J. Phys. Chem. B, **109**, 9082–9095, 2005.
32. D. Riccardi, G. Li, and Q. Cui, *The importance of van der Waals interactions in QM/MM simulations*, J. Phys. Chem. B, **108**, 6467–6478, 2004.
33. A. Laio, J. VanderVondele, and U. Rothlisberger, *A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations*, j. Chem. Phys., **116**, 6941–6947, 2002.
34. M. Wanko, M. Hoffmann, T. Frauenheim, and M. Elstner, *Effect of polarization on the opsin shift in rhodopsins. 1. A combined QM/QM/MM model for bacteriorhodopsin and pharaonis sensory rhodopsin II*, J. Phys. Chem. B, **112**, 11462–11467, 2008.
35. M. Wanko, M. Hoffmann, J. Frahmcke, T. Frauenheim, and M. Elstner, *Effect of polarization on the opsin shift in rhodopsins. 2. empirical polarization models for proteins*, J. Phys. Chem. B, **112**, 11468–11478, 2008.
36. Daan Frenkel and Berend Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, San Diego, London, 2002.
37. D. J. Wales, *Energy Landscapes*, Cambridge University Press, 2003.
38. T. Siomonson, *Computational biochemistry and biophysics*, Marcel Dekker, Inc., 2001.
39. S. Fischer and M. Karplus, *Conjugate Peak Refinement : an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom.*, Chem. Phys. Lett., **194**, 252–261, 1992.
40. G. Henkelman, B. P. Uberuaga, and H. Jónsson, *Climbing image nudged elastic band*

method for finding saddle points and minimum energy paths, J. Chem. Phys., **113**, 9901–9904, 2000.
41. G. Henkelman and H. Jónsson, *A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives*, J. Chem. Phys., **111**, 7010–7022, 1999.
42. A. Bondar, S. Fischer, J. C. Smith, M. Elstner, and S. Suhai, *Key role of electrostatic interactions in bacteriorhodopsin proton transfer*, Journal of the American Chemical Society, **126**, 14668–14677, 2004.
43. M. Klahn, S. Braun-Sand, E. Rosta, and A. Warshel, *On possible pitfalls in ab initio quantum mechanics/molecular mechanics minimization approaches for studies of enzymatic reactions*, J. Phys. Chem B, **109**, 15645, 2005.
44. Y. K. Zhang, H. Y. Liu, and W. T. Yang, *Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface*, J. Chem. Phys., **112**, 3483–3492, 2000.
45. H. Hu, Z. Y. Lu, and W. T. Yang, *QM/MM minimum free-energy path: Methodology and application to triosephosphate isomerase*, J. Chem. Theo. Comp., **3**, 390–406, 2007.
46. H. Hu and W. T. Yang, *Free Energies of Chemical Reactions in Solution and in Enzymes with Ab Initio Quantum Mechanics/Molecular Mechanics Methods*, Annu. Rev. Phys. Chem., **59**, 573–601, 2008.
47. G.M. Torrie and J.P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*, J. Comp. Phys., **23**, 187–199, 1977.
48. A. Laio and M. Parrinello, *Escaping free energy minima*, Proc. Nat. Acad. Sci. USA, **99**, 12562–12566, 2002.
49. A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello, *Assessing the accuracy of metadynamics*, J. Phys. Chem. B, **109**, 6714–6721, 2005.
50. A. Barducci, G. Bussi, and M. Parrinello, *Well-tempered metadynamics: A smoothly converging and tunable free-energy method*, Phys. Rev. Lett., **100**, 020603, 2008.
51. D. H. Min, Y. S. Liu, I. Carbone, and W. Yang, *On the convergence improvement in the metadynamics simulations: A Wang-Landau recursion approach*, J. Chem. Phys., **126**, 194104, 2007.
52. H. Li, D. Min, Y. Liu, and W. Yang, *Essential energy space random walk via energy space metadynamics method to accelerate molecular dynamics simulations*, J. Chem. Phys., **127**, 094101, 2007.
53. Y. Q. Gao, *An integrate-over-temperature approach for enhanced sampling*, J. Chem. Phys., **128**, 064105, 2008.
54. D. Hamelberg, J. Mongan, and J. A. Mccammon, *Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules*, J. Chem. Phys., **120**, 11919–11929, 2004.
55. P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Transition path sampling: Throwing ropes over rough mountain passes, in the dark*, Annu. Rev. Phys. Chem., **53**, 291–318, 2002.
56. R. Crehuet and M. J. Field, *A transition path sampling study of the reaction catalyzed by the enzyme chorismate mutase*, J. Phys. Chem. B, **111**, 5708–5718, 2007.
57. S. Saen-oon, S. Quaytman-Machleder, V. L. Schramm, and S. D. Schwartz, *Atomic*

detail of chemical transformation at the transition state of an enzymatic reaction, Proc. Natl. Acad. Sci. USA, **105**, 16543–16548, 2008.
58. M. Elstner, T. Frauenheim, J. McKelvey, and G. Seifert, *Density functional tight binding: Contributions from the American chemical society symposium*, J. Phys. Chem. A, **111**, 5607–5608, 2007.
59. W. Thiel, *Perspectives on semiempirical molecular orbital theory*, Adv. Chem. Phys., **93**, 703–757, 1996.
60. I. Rossi and D. G. Truhlar, *Parameterization of NDDO wavefunctions using genetic algorithm*, Chem. Phys. Lett., **233**, 231–236, 1995.
61. Q. Cui and M. Karplus, *QM/MM Studies of the Triosephosphate Isomerase (TIM) Catalyzed Reactions: Verification of Methodology and Analysis of the Reaction Mechanisms*, J. Phys. Chem. B, **106**, 1768–1798, 2002.
62. K. Nam, Q. Cui, J. Gao, and D. M. York, *A specific reaction parameterization for the AM1/d Hamiltonian for transphosphorylation reactions*, J. Chem. Theo. Comp., **3**, 486–504, 2007.
63. Yang Yang, Haibo Yu, Darrin M. York, Marcus Elstner, and Qiang Cui, *Description of phosphate hydrolysis reactions with the Self-Consistent-Charge Tight-Binding-Density-Functional (SCC-DFTB) theory 1. Parameterization*, J. Chem. Theo. Comp., **In press**, 2008.
64. M. P. Repasky, J. Chandrasekhar, and W. L. Jorgensen, *PDDG/PM3 and PDDG/MNDO: Improved semiempirical methods*, J. Comp. Chem., **23**, 1601–1622, 2002.
65. Porezag, D., Frauenheim, T., Köhler, T., Seifert, G., and Kaschner, R., *construction of tight-binding-like potentials on the basis of density functional theory - application to carbon*, Phys. Rev. B, **51**, 12947–12957, 1995.
66. M. Elstner, D. Porezag, G. Jungnickel, J. Elstner, M. Haugk, Th. Frauenheim, S. Suhai, and G. Seifert, *Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties*, Phys. Rev. B, **58**, 7260–7268, 1998.
67. M. Elstner, *SCC-DFTB: What is the proper degree of self-consistency*, Journal of Physical Chemistry A, **111**, 5614–5621, 2007.
68. Yang, Y., Yu, H., York, D., Cui, Q., and Elstner, M., *Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method: Third-Order Expansion of the Density Functional Theory Total Energy and Introduction of a Modified Effective Coulomb Interaction*, J. Phys. Chem. A, **111**, 10861–10873, 2007.
69. T. Kruger, M. Elstner, P. Schiffels, and Th. Frauenheim, *Validation of the density functional based tight-binding approximation method for the calculation of reaction energies and other data*, J. Chem. Phys., **122**, 114110, 2005.
70. K. W. Sattelmeyer, J. Tirado-Rives, and W. L. Jorgensen, *Comparison of SCC-DFTB and NDDO-based semiempirical molecular orbital methods for organic molecules*, Journal of Physical Chemistry A, **110**, 13551–13559, 2006.
71. N. Otte, M. Scholten, and W. Thiel, *Looking at self-consistent-charge density functional tight binding from a semiempirical perspective*, Journal of Physical Chemistry A, **111**, 5751–5755, 2007.
72. A. Bondar, S. Suhai, S. Fischer, J. C. Smith, and M. Elstner, *Suppression of the back proton-transfer from Asp85 to the retinal Schiff base in bacteriorhodopsin: A theoretical analysis of structural elements*, Journal of Structural Biology, **157**, 454–469,

2007.
73. M. Wano, M. Hoffmann, P. Strodel, A. Koslowski, W. Thiel, F. Neese, T. Frauenheim, and M. Elstner, *Calculating absorption shifts for retinal proteins: Computational challenges*, J. Phys. Chem. B, **109**, 3606–3615, 2005.
74. M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, and E. Kaxiras, *Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density functional-theory based treatment*, Journal of Chemical Physics, **114**, 5149–5155, 2001.
75. A. Dreuw, J. L. Weisman, and M. Head-Gordon, *Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange*, Journal of Chemical Physics, **119**, 2943–2946, 2003.
76. M. Wanko, M. Garavelli, F. Bernardi, T. A. Niehaus, T. Frauenheim, and M. Elstner, *A global investigation of excited state surfaces within time-dependent density-functional response theory*, Journal of Chemical Physics, **120**, 1674–1692, 2004.
77. F. Claeyssens, J. N. Harvey, F. R. Manby, R. A. Mata, A. J. Mulholland, K. E. Ranaghan, M. Schütz, S. Thiel, W. Thiel, and H. J. Werner, *High-accuracy computation of reaction barriers in enzymes*, Angew. Chim. Intl. Ed., **45**, 6856–6859, 2006.
78. J. A. Wagoner and N. A. Baker, *Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms*, Proc. Nat. Acad. Sci. USA, **103**, 8331–8336, 2006.
79. W. Im, S. Berneche, and B. Roux, *Generalized solvent boundary potential for computer simulations*, J. Chem. Phys., **114**, 2924–2937, 2001.
80. D. Riccardi and Q. Cui, *pK_a analysis for the zinc-bound water in Human Carbonic Anhydrase II: benchmark for "multi-scale" QM/MM simulations and mechanistic implications*, J. Phys. Chem. A, **111**, 5703–5711, 2007.
81. A. R. Dinner, X. Lopez, and M. Karplus, *A charge-scaling method to treat solvent in QM/MM simulations*, Theoretical Chemistry Accounts, **109**, 118–124, 2003.
82. G. Li, X. Zhang, and Q. Cui, *Free Energy Perturbation Calculations with Combined QM/MM Potentials Complications, Simplifications, and Applications to Redox Potential Calculations*, J. Phys. Chem. B, **107**, 8643–8653, 2003.
83. W. G. Han, K. J. Jalkanen, M. Elstner, and S. Suhai, *Theoretical study of aqueous N-acetyl-L-alanine N'-methylamide: Structures and Raman, VCD, and ROA spectra*, Journal of Physical Chemistry B, **102**, 2587–2602, 1998.
84. H. Hu, M. Elstner, and J. Hermans, *Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution*, Proteins: Structure, Function and Genetics, **50**, 451–463, 2003.
85. H. Liu, M. Elstner, E. Kaxiras, T. Frauenheim, J. Hermans, and W. Yang, *Quantum mechanics simulation of protein dynamics on long timescale*, Proteins: Structure, Function and Genetics, **44**, 484–489, 2001.
86. J. M. J. Swanson, C. M. Maupin, H. Chen, M. K. Petersen, J. Xu, Y. Wu, and G. A. Voth, *Proton solvation and transport in aqueous and biomolecular systems: insights from computer simulations*, J. Phys. Chem. B, **111**, 4300–4314, 2007.
87. A. Bondar, M. Elstner, S. Suhai, J. C. Smith, and S. Fischer, *Mechanism of primary proton transfer in bacteriorhodopsin*, Structure, **12**, 1281–1288, 2004.

88. M. Elstner, *The SCC-DFTB method and its application to biological systems*, Theor. Chem. Acc., **116**, 316–325, 2006.
89. K. Håkansson, M. Carlsson, L. A. Svensson, and A. Liljas, *Structure of native and apo carbonic anhydrase II and structure of some its anion-ligand complexes*, J. Mol. Biol., **227**, 1192–1204, 1992.
90. D. N. Silverman, *Proton transfer in carbonic anhydrase measured by equilibrium isotope exchange*, Methods in Enzymology, **249**, 479–503, 1995.
91. D. Riccardi, P. König, X. Prat-Resina, H. Yu, M. Elstner, T. Frauenheim, and Q. Cui, *"Proton holes" in long-range proton transfer reactions in solution and enzymes: A theoretical analysis*, J. Am. Chem. Soc., **128**, 16302–16311, 2006.
92. D. Riccardi, P. Koenig, H. Guo, and Q. Cui, *Proton Transfer in Carbonic Anhydrase Is Controlled by Electrostatics Rather than the Orientation of the Acceptor*, Biochem., **47**, 2369–2378, 2008.
93. P. H. Konig, N. Ghosh, M. Hoffmann, M. Elstner, E. Tajkhorshid, Th Frauenheim, and Q. Cui, *Toward theoretical analyis of long-range proton transfer kinetics in biomolecular pumps*, Journal of Physical Chemistry A, **110**, 548–563, 2006.
94. B. Roux, *Influence of the membrane potential on the free energy of an intrinsic protein*, Biophys. J., **73**, 2980–2989, 1997.
95. N. Ghosh, X. Prat-Resina, M. Gunner, and Q. Cui, *Microscopic pK_a analysis of Glu 286 in Cytochrome c Oxidase (Rhodobacter sphaeroids): towards a calibrated molecular model*, Biochem., **Submitted**, 2008.
96. M. Kato, A. V. Pisliakov, and A. Warshel, *The barrier for proton transport in Aquaporins as a challenge for electrostatic models: The role of protein relaxation in mutational calculations*, Proteins: Struct. Funct. Bioinfor., **64**, 829–844, 2006.
97. P. Phatak, N. Ghosh, H. Yu, M. Elstner, and Q. Cui, *Amino acids with an intermolecular proton bond as proton storage site in bacteriorhodopsin*, Proc. Acad. Natl. Sci. U.S.A., **In press**, 2008.
98. F. Garczarek, L. S. Brown, J. K. Lanyi, and K. Gerwert, *Proton binding within a membrane protein by a protonated water cluster*, Proc. Acad. Natl. Sci. U.S.A., **102**, 3633–3638, 2005.
99. F. Garczarek and K. Gerwert, *Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy*, Nature, **439**, 109–112, 2006.
100. M. Hoffmann, M. Wanko, P. Strodel, P. H. Konig, T. Frauenheim, K. Schulten, W. Thiel, E. Tajkhorshid, and M. Elstner, *Color tuning in rhodopsins: The mechanism for the spectral shift between bacteriorhodopsin and sensory rhodopsin II*, J. Am. Chem. Soc., **128**, 10808–10818, 2006.